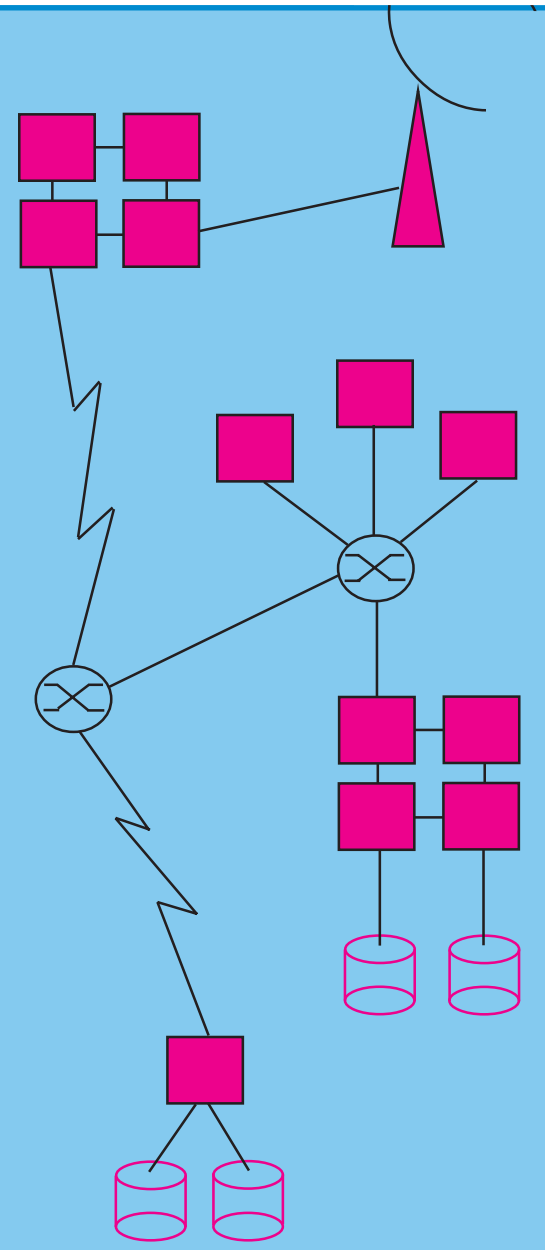
Globus, Nexus, and NT

Ian Foster and Steven Tuecke
Argonne National laboratory

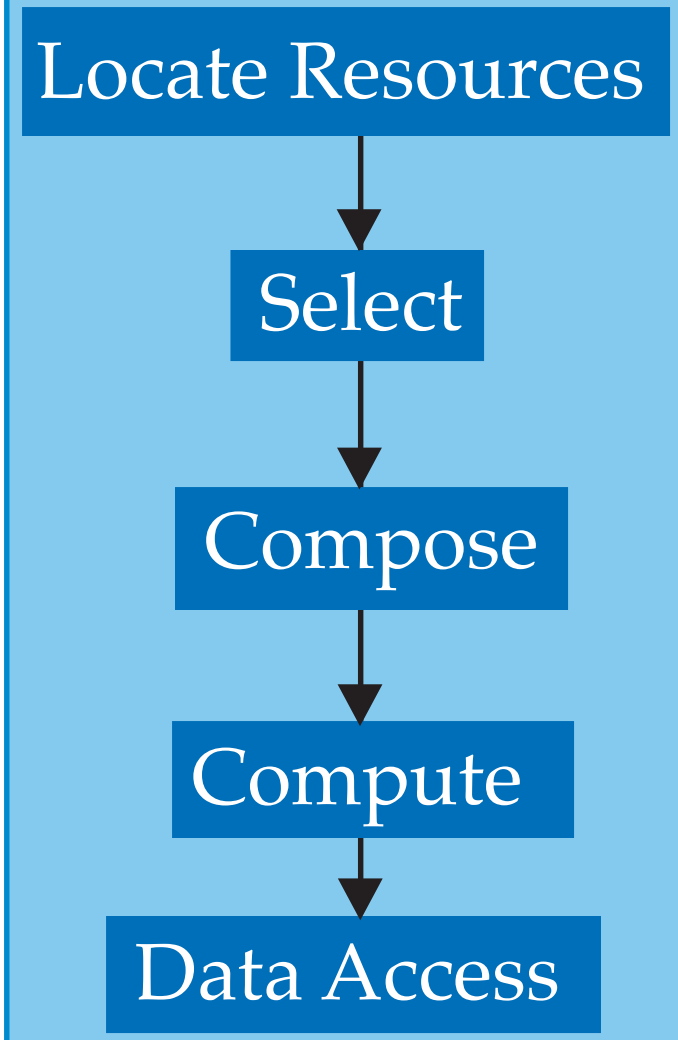
http://www.globus.org/

Networked Virtual Supercomputers

- Assemble distributed resources …
  - High-end computers
  - Information sources
  - Scientific instruments, etc.
- … and apply to challenging problems
  - Smart instruments
  - Collaborative engineering
  - Data mining
- What role can (NT) clusters play?



Technical Requirements



Resource naming & location
Scalable authentication & authorization

Identifying and selecting bitways
Scheduling computers and networks

Integrating resources into computations
Configuration

Multiple programming models
Masking latency, maximizing bandwidth

Uniform and efficient access

Software Challenges

- Development of required services
  - Many promising pieces, much missing functionality
- Integration of existing and new services
  - Requires common low-level mechanisms
- Meeting demanding performance requirements
  - Configuration and adaptation are critical
- Evaluation and intercomparison of solutions
  - Requires instrumentation, testbeds, benchmarks

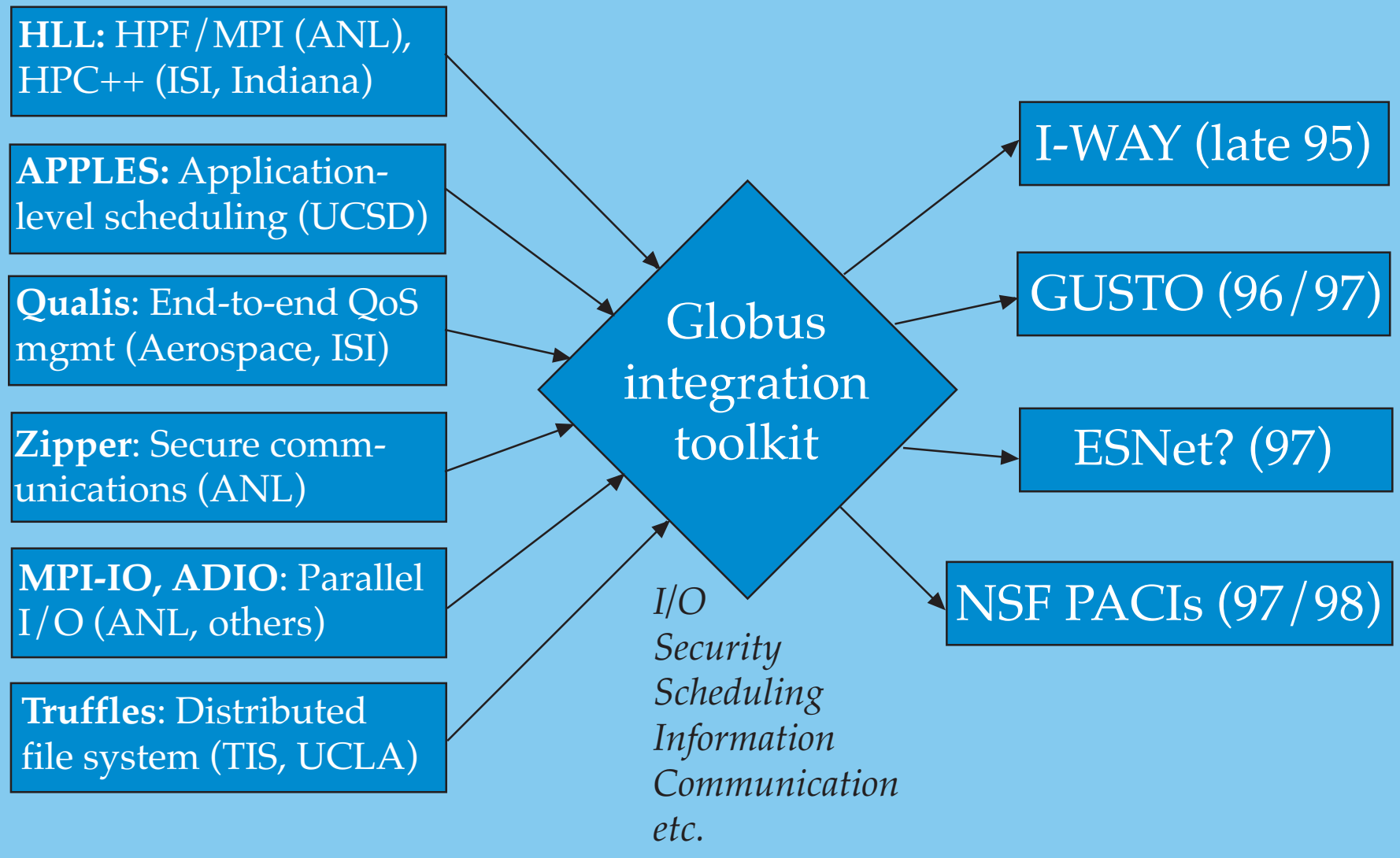
The Configuration Problem

- Heterogeneity demands many choices
  - Selection of resources and networks
  - Configuration of networks (QoS) & devices
  - Communication protocols, security, …
  - Configuration of computation
- End-to-end management of complex systems
  - Manageable complexity
  - Configuration without omniscience
- Distinguishes us from distributed computing

Components of a Solution

- Toolkit providing low-level mechanisms
  - Encapsulate device-specific mechanisms
  - Control, instrumentation, notification interfaces
  - Permit high-level specification of policy
- Universal and uniform access to up-to-date information on system structure and state
- Autoconfiguration and adaptation mechanisms for individual components
- Integration of diverse higher-level services

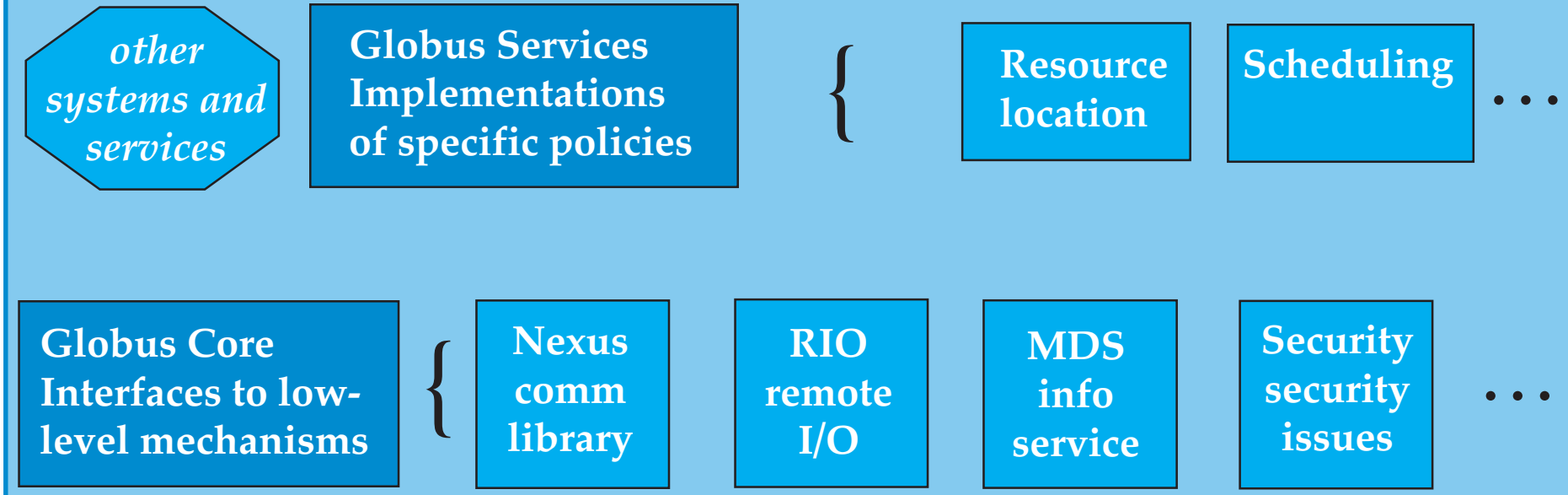
Globus Approach



Globus Structure

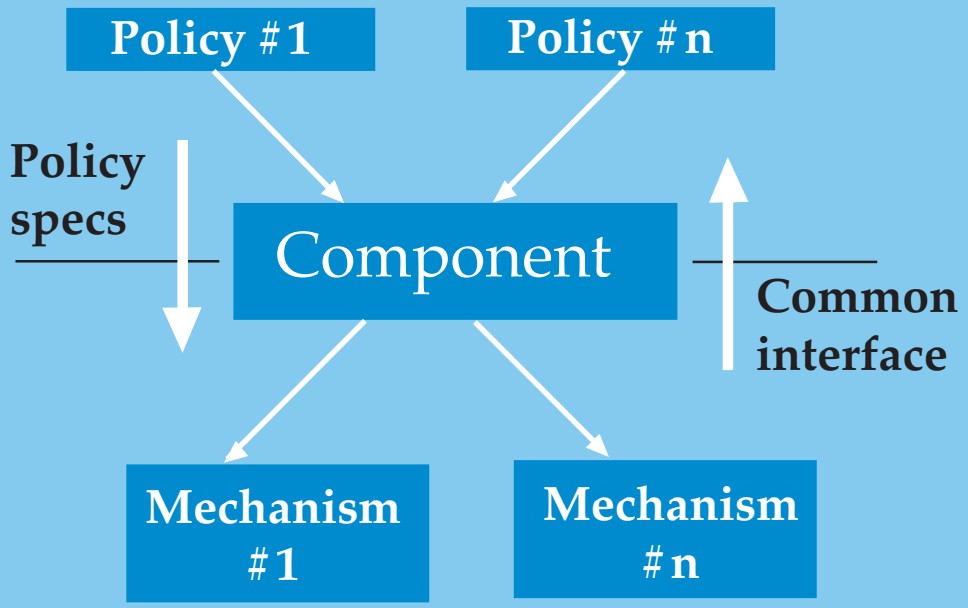
HPC++
MPI
Condor
etc.

Legion
AppLeS
Nimrod



Characteristics of a Globus Infrastructure Component

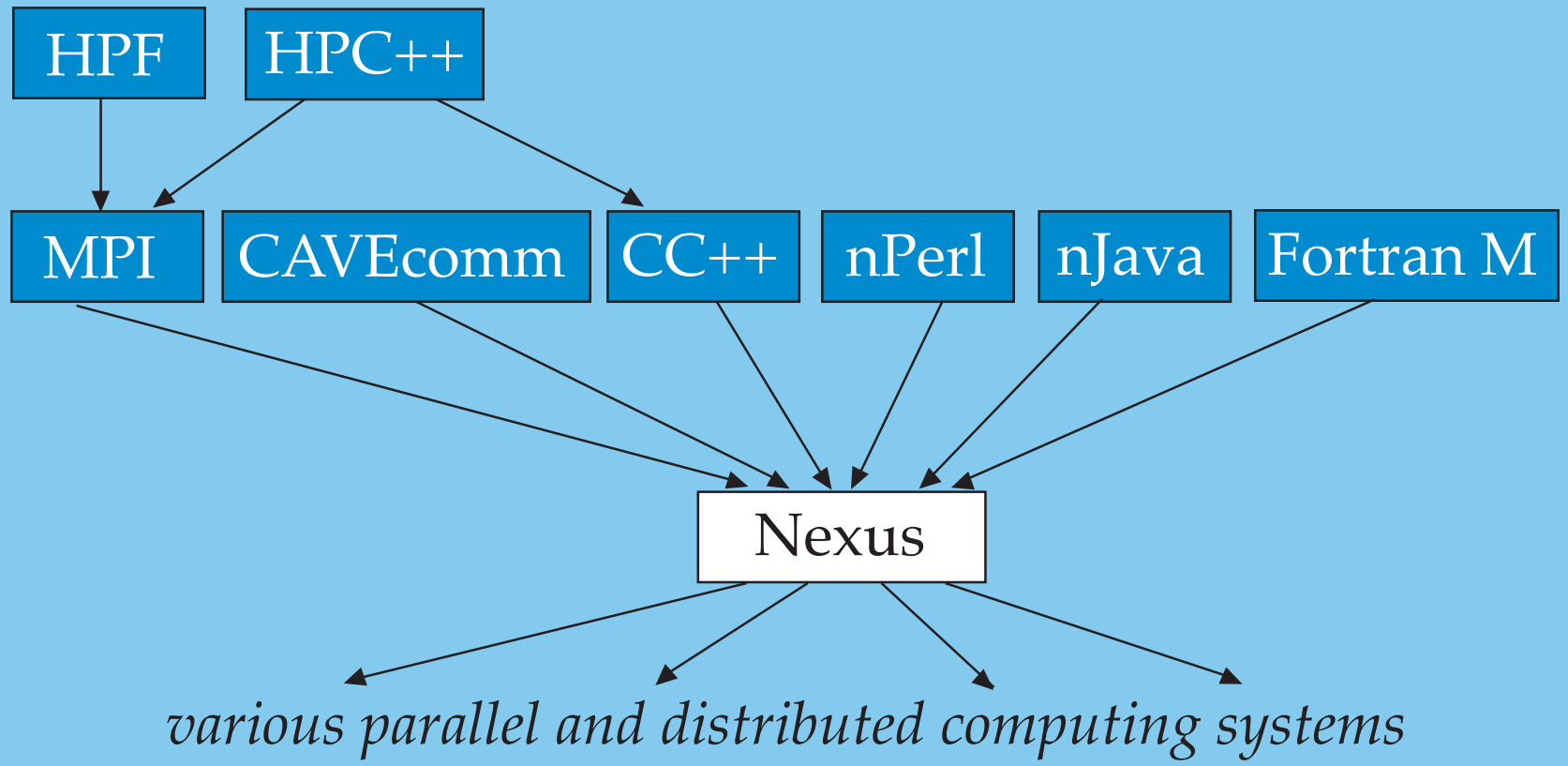
- Provide common interface to low-level mechanisms
- Permit higher-level components (tools or users) to specify "policy"
- Permit high-performance implementations
- Use information service for configuration



Communications Issues

- Challenges
  - Heterogeneous devices and networks
  - Need for multiple communication methods
  - Support irregular, dynamic program structures
- Approach
  - Common communication infrastructure (Nexus)
  - Separate treatment of policy and mechanism
  - Automatic/manual management of method choices
  - Support wide range of high-level tools

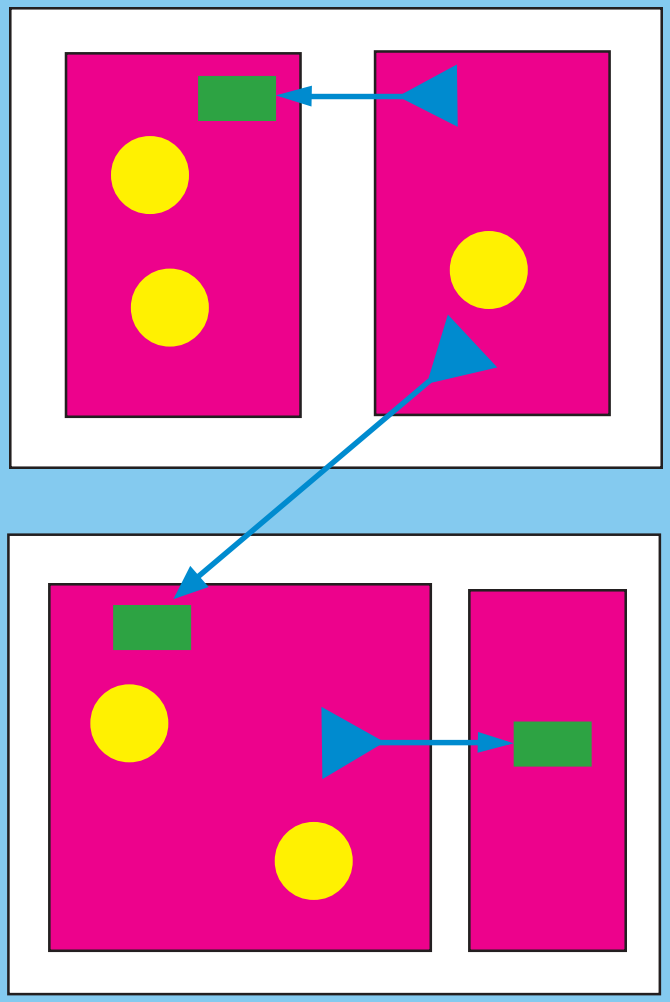
Nexus Applications



Nexus Mechanisms

- **Node**: locus of computation
- **Context** : address space
  (virtual processor)
- **Thread** : thread of control
- **Communication link** :
  uniform naming
- **Remote service request** :
  remote invocation

All entities can be dynamically created and destroyed; links can be migrated



Communication Links

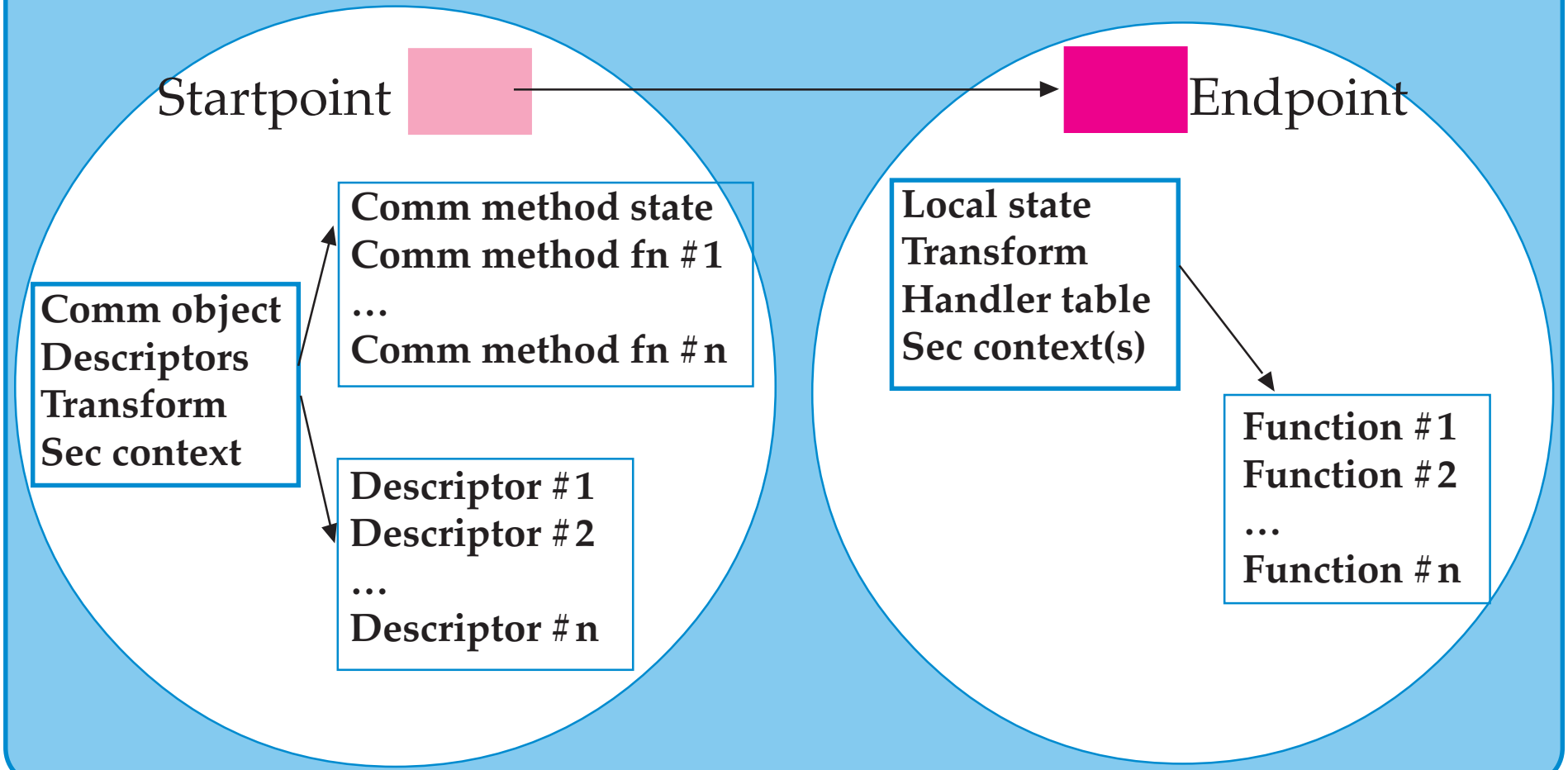
init_rsr(…)
put_int(…)
put_float(…)
send–rst(…)

4-12	4	1	4
endpoint addr	handler id	mach type	size

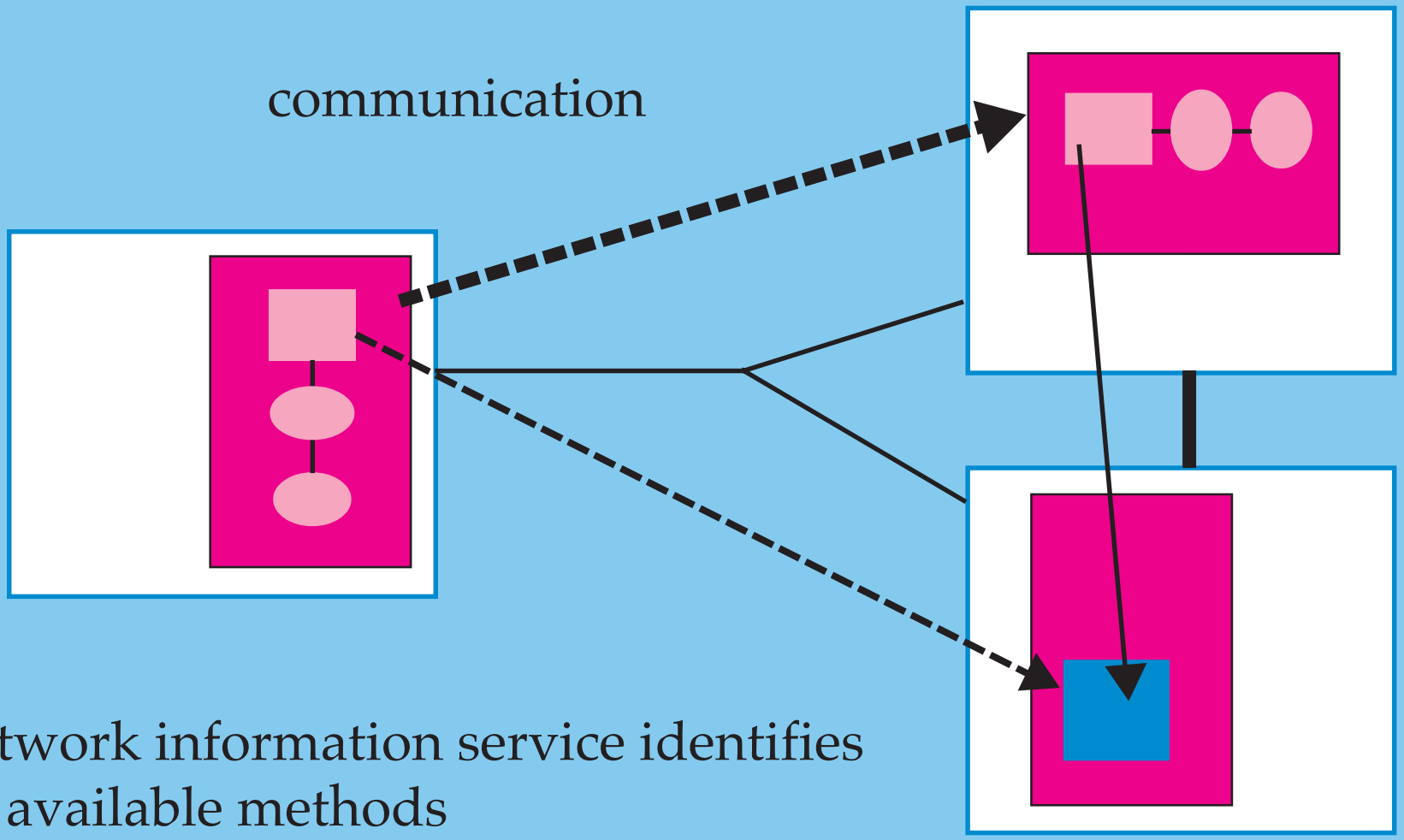
message header

bytes

handler {
  get_int(…)
  get_float(…)
}



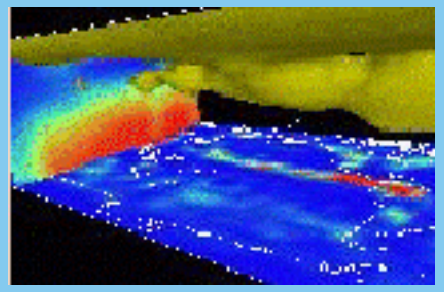
Communication Method Selection



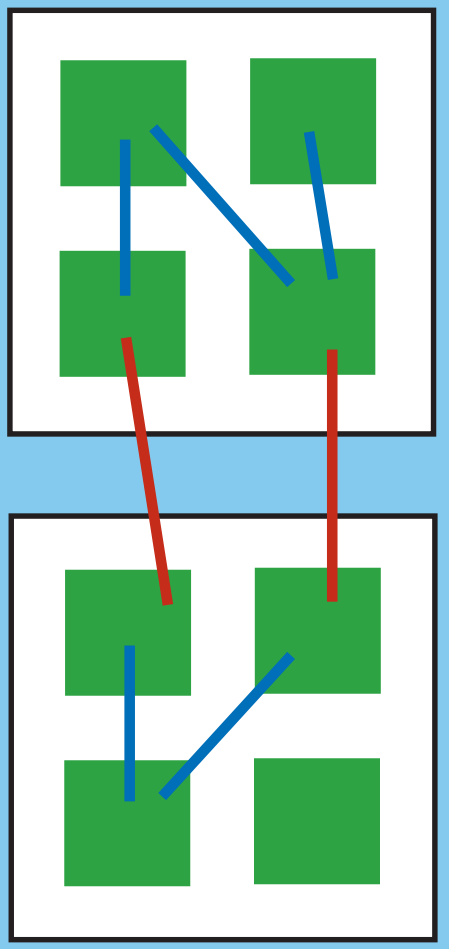
Network information service identifies
available methods
Default rule: select “fastest” method
Programmer selection also supported

Multimethod Communication: Heterogeneous Systems

- Select communication method according to destination
- Example: coupled climate model
  - Ocean, atmos. on separate computers
  - Written entirely in MPI



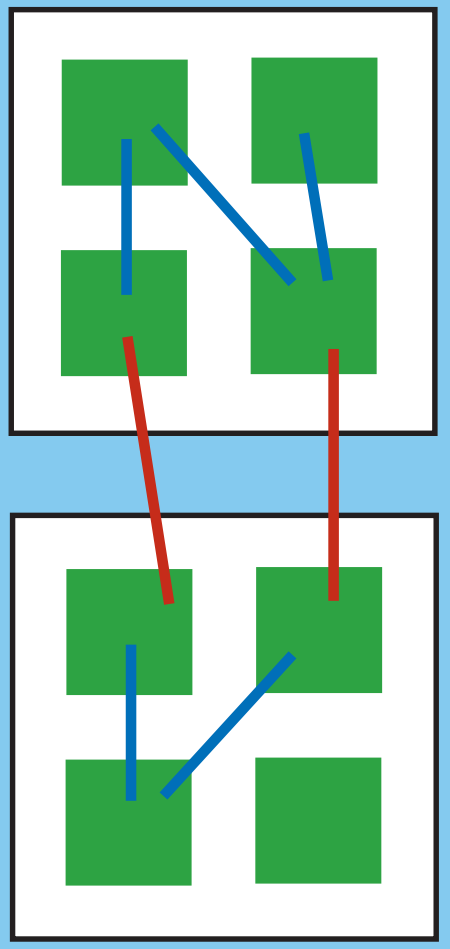
Scenario	Partitions	Total
All MPL	1	560
TCP+MPL	2	574
All TCP	2	854



Multimethod Communication: Security

- Associate security mechanisms with communication link
- Select mechanism according to destination
- Example: coupled climate model

Security	TCP time (secs/day)	MPL time (secs/day)
None	854	574
Coupler	897	590
All	1459	1187



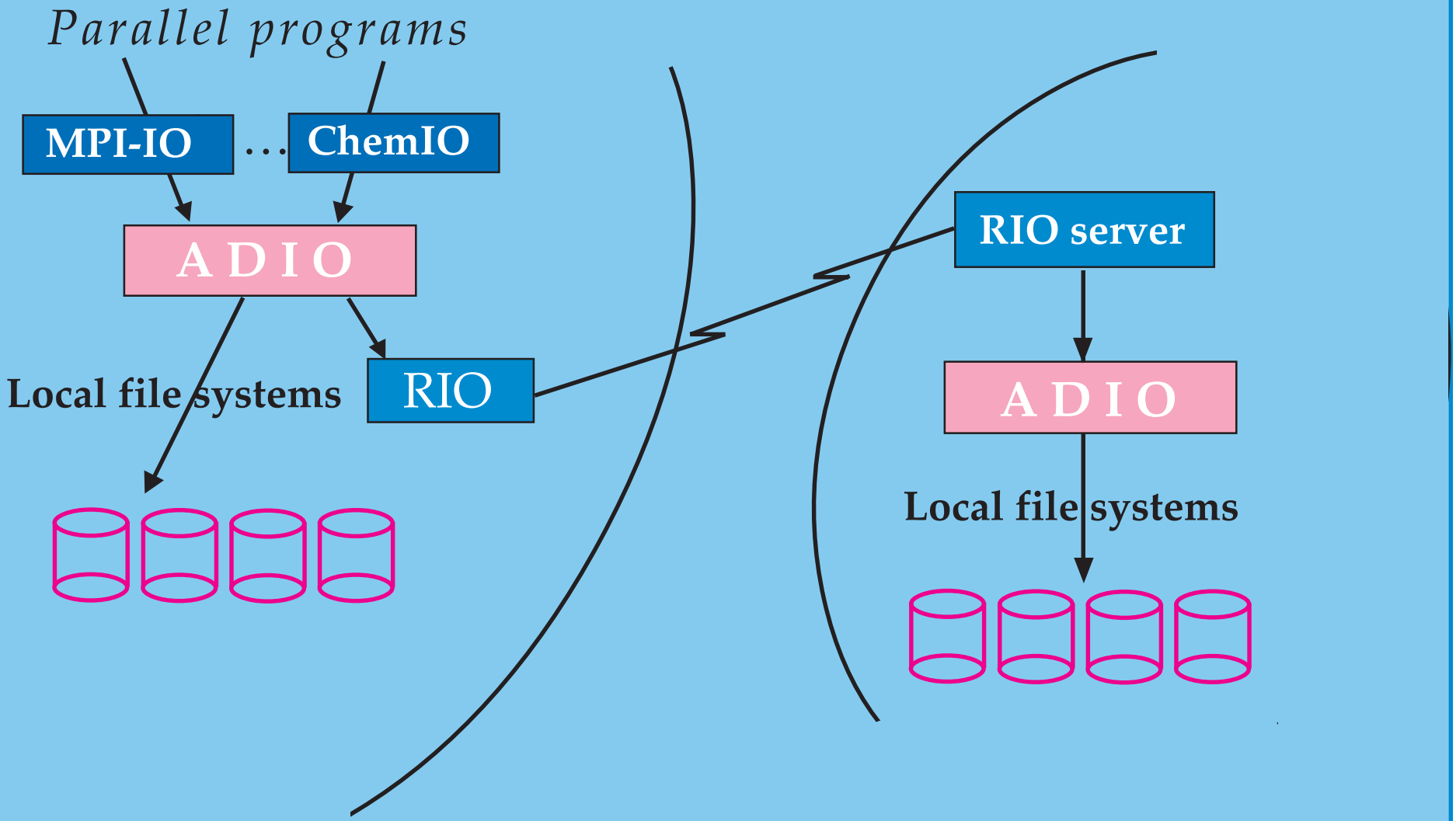
= Unsecure

= Secure (DES/ECB)

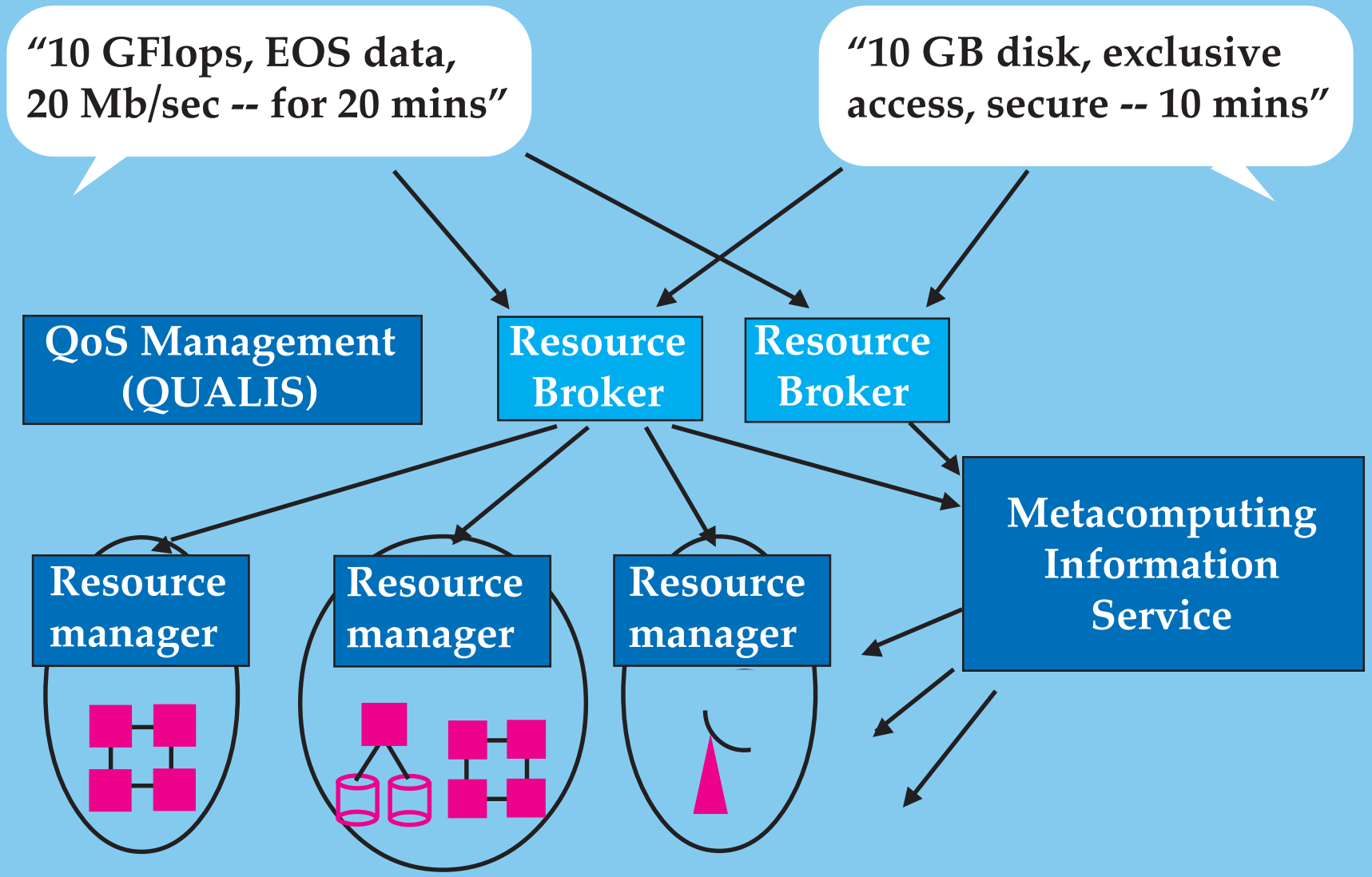
Component Example: RIO

- Purpose: Support high-speed access from (parallel) programs to remote (parallel) files
- Approach: Use ADIO as interface, implement network I/O device support
- Build on Globus communication and information services to guide configuration
- A building block for I/O libraries (e.g., ChemIO, MPI-IO) and file systems

RIO contd.


Parallel programs
MPI-IO
...
ChemIO
A D I O
Local file systems
RIO
RIO server
A D I O
Local file systems

Resource Management Components



Location and Allocation Services

- Low-level services: Resource Manager
  - Determine scheduling characteristics of resource
  - Enquire about availabilityof specified resources
  - Allocate resources via queue or reservation
  - Support variety of resources and brokers
- Higher-level services: Resource Broker
  - Translate user requirements into resource allocations

Next Steps

- Development/deployment
  - MDS, location/allocation, communication, process management, security, NT
  - Interfaces to AppLeS, LSF, Condor, Nimrod, etc.
- Research
  - Higher-level services & resource aware applications
  - Scalability and performance issues
- Applications
  - Smart instruments, collaborative environments, etc.

Deployment

- GUSTO : A metacomputing testbed
  - Testbed for Globus components and friendly applications
  - April 97 : 4 - 6 sites within U.S.
  - October 97 : 10 - 12 sites (some international)
- NSF supercomputer centers: NCSA, SDSC
- DOE sites : Center for Computational Science and Technology, others

Cluster Issues

- Is an NT cluster a reasonable platform for network computing?
- Integrating non-IP cluster communication
- Scheduling integration
- Information services required for clusters
- Clusters with heterogeneous nodes

Globus Summary

- Defined and implemented basic services to support "network supercomputing"
  - Comms., info., security, I/O, sched., etc.
- Demonstrated autoconfiguration in network computing systems: MPI, etc.
- Demonstrated utility in large-scale testbeds
- Next steps: adaptation, integration, applications, larger testbeds